

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 15:43:11 ; Search time 60 Seconds
(without alignments)
2283.926 Million cell updates/sec

Title: US-09-769-736-72
Perfect score: 2621
Sequence: 1 MFFLAYNAFTAIGVSIPHCN.....QSFHKQLIQLSNKYSVPES 485

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: Geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2621	100.0	485	3 AAY91310	Aay91310 Group B S
2	1885	71.9	379	5 ABP26206	Abp26206 Streptoco
3	563	21.5	816	3 AAB12756	Aab12756 Streptoco
4	563	21.5	820	3 AAB12755	Aab12755 Streptoco
5	563	21.5	838	3 AAB01466	Aab01466 Recombina
6	563	21.5	839	6 ABU01418	Abu01418 S. pneumo
7	562	21.4	811	3 AAB12761	Aab12761 Streptoco
8	562	21.4	811	3 AAB12762	Aab12762 Streptoco
9	562	21.4	811	3 AAB12760	Aab12760 Streptoco
10	562	21.4	816	3 AAB12758	Aab12758 Streptoco
11	562	21.4	816	3 AAB12757	Aab12757 Streptoco
12	558	21.3	568	3 AAB12741	Aab12741 Streptoco
13	558	21.3	568	5 AAU84088	Aau84088 Truncated
14	558	21.3	568	7 ABM18821	Abm18821 S. pneumo
15	558	21.3	612	5 AAU84031	Aau84031 Truncated
16	558	21.3	612	7 ABM18832	Abm18832 S. pneumo
17	558	21.3	819	3 AAB12754	Aab12754 Streptoco
18	558	21.3	819	3 AAB12740	Aab12740 Streptoco
19	558	21.3	819	5 AAU84087	Aau84087 Truncated
20	558	21.3	819	7 ABM18820	Abm18820 S. pneumo
21	558	21.3	838	3 AAB12720	Aab12720 Streptoco
22	558	21.3	838	5 AAU75934	Aau75934 Streptoco
23	558	21.3	838	7 ABM18798	Abm18798 S. pneumo
24	558	21.3	1139	5 AAU84055	Aau84055 S. pneumo
25	558	21.3	1139	7 ABM18839	Abm18839 S. pneumo

26	558	21.3	1378	5 AAU84053	Aau84053 S. pneumo
27	558	21.3	1378	7 ABM18837	Abm18837 S. pneumo
28	552	21.1	1238	5 AAU84056	Aau84056 S. pneumo
29	552	21.1	1238	7 ABM18840	Abm18840 S. pneumo
30	547	20.9	466	5 ABP26208	Abp26208 Streptoco
31	546.5	20.9	819	6 ABU01597	Abu01597 S. pneumo
32	546.5	20.9	834	3 AAB12759	Aab12759 Streptoco
33	546	20.8	805	3 AAB12764	Aab12764 Streptoco
34	545.5	20.8	721	2 AAY05753	Aay05753 Streptoco
35	545.5	20.8	819	3 AAB01469	Aab01469 Recombina
36	545.5	20.8	826	3 AAY91939	Aay91939 S. pneumo
37	545.5	20.8	827	3 AAY81662	Aay81662 Streptoco
38	543	20.7	555	3 AAB12734	Aab12734 Streptoco
39	543	20.7	555	5 AAU84047	Aau84047 Truncated
40	543	20.7	555	7 ABM18814	Abm18814 S. pneumo
41	543	20.7	613	3 AAB12730	Aab12730 Streptoco
42	543	20.7	613	5 AAU84029	Aau84029 Truncated
43	543	20.7	613	7 ABM18810	Abm18810 S. pneumo
44	543	20.7	807	3 AAB12765	Aab12765 Streptoco
45	543	20.7	821	3 AAB12766	Aab12766 Streptoco

ALIGNMENTS

RESULT 1
AAY91310
ID AAY91310 standard; protein; 485 AA.
XX
AC AAY91310;
XX
DT 30-MAY-2000 (first entry)
XX
DE Group B Streptococcus protein sequence SEQ ID NO:38.
XX
KW Group B Streptococcus; Streptococcus agalactiae; protein antigen;
KW vaccine; screening; immunogen; detection; diagnosis; infection; antibody;
KW affibody; antibacterial.
XX
OS Streptococcus agalactiae.
XX
PN WO200006736-A2.
XX
PD 10-FEB-2000.
XX
PF 27-JUL-1999; 99WO-GB002444.
XX
PR 27-JUL-1998; 98GB-00016335.
PR 19-MAR-1999; 99US-0125163P.
XX
(MICR-) MICROBIAL TECHNICS LTD.
XX
PI Le Page RWF, Wells JM, Hanniffy SB;
XX
DR WPI; 2000-195299/17.
XX
PT New Group B Streptococcus protein, useful as vaccine, for diagnosis of
Streptococcal infections and for screening of antibodies or affibodies.
XX
PS Claim 1; Fig 1; 123pp; English.
XX
CC AAA05803 to AAA05872 encode proteins, polypeptides and peptides (given in
CC AAY91275 to AAY91343) isolated from Group B Streptococcus (GBS), also
CC known as Streptococcus agalactiae. The GBS polynucleotides and
CC polypeptides have antibacterial activity. Immunogenic compositions
CC comprising GBS polynucleotides or polypeptides can be used as vaccines
CC and for the treatment or prophylaxis of GBS infection. The
CC polynucleotides and polypeptides can also be used in the detection of GBS
CC and for screening DNA encoding bacterial cell envelope associated or
CC secreted antigens in gram positive bacteria. AAA05873 to AAA05941
CC represent primers used in the exemplification of the present invention
XX
SQ Sequence 485 AA;

Query Match 100.0%; Score 2621; DB 3; Length 485;
Best Local Similarity 100.0%; Pred. No. 1.4e-201;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEFLAYNAFTAIGVSIPGHNHFHFIHYKDMSPLEATRMVAEHRGHIDALGKDKSTK 60
DB 1 MEFLAYNAFTAIGVSIPGHNHFHFIHYKDMSPLEATRMVAEHRGHIDALGKDKSTK 60

QY 61 PKHISHEPNKEPHTTEEEHHAHTPKDQKKGKPNQIVYSAQIEEAKKAGYTTSDGYIFD 120
DB 61 PKHISHEPNKEPHTTEEEHHAHTPKDQKKGKPNQIVYSAQIEEAKKAGYTTSDGYIFD 120

QY 121 AKDIKKTGTGYVIPHMTHEHWPVKDLSESELKAAQEFSLGKSEANQDKPKTKGTAQEI 180
DB 121 AKDIKKTGTGYVIPHMTHEHWPVKDLSESELKAAQEFSLGKSEANQDKPKTKGTAQEI 180

QY 181 YEALPKAIVKPEDLLFGIAQATDYKNGTFVIPHKDHVHYVELKWFDEKDLLADSDKY 240
DB 181 YEALPKAIVKPEDLLFGIAQATDYKNGTFVIPHKDHVHYVELKWFDEKDLLADSDKY 240

QY 241 SLEDYLATAKYMMHPEKRPKEGVGKDAEIIYKEKSNKADKPSAPTDNKSTNSSDKN 300
DB 241 SLEDYLATAKYMMHPEKRPKEGVGKDAEIIYKEKSNKADKPSAPTDNKSTNSSDKN 300

QY 301 LSAAEVFKQAKPEKIVPLDKIAAHMAYAVGFEDDQILVPHHDHYHVPMAWFDKGLWKA 360
DB 301 LSAAEVFKQAKPEKIVPLDKIAAHMAYAVGFEDDQILVPHHDHYHVPMAWFDKGLWKA 360

QY 361 PEGYTLQOLFSTIKYMEHPNELPKKEGWGHDSHNKSNKDNKAKNYAPDEEEDSGKV 420
DB 361 PEGYTLQOLFSTIKYMEHPNELPKKEGWGHDSHNKSNKDNKAKNYAPDEEEDSGKV 420

QY 421 THNYGYDVNKGSDDEEPEKQDESELDEYELGMAQNAKKGMDRQSFQKLIQLSNKYS 480
DB 421 THNYGYDVNKGSDDEEPEKQDESELDEYELGMAQNAKKGMDRQSFQKLIQLSNKYS 480

QY 481 VSFES 485
DB 481 VSFES 485

RESULT 2
ID ABP26206 standard; protein; 379 AA.
XX
AC ABP26206;
XX
DT 02-JUL-2002 (first entry)
XX
DE Streptococcus polypeptide SEQ ID NO 1588.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS Streptococcus agalactiae.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB004789.
XX
PR 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Massignani V, Margarit Y Rosi, Grandi G, Fraser C,
PI Tettelin H;

XX WPI; 2002-352536/38.
DR N-PSDB; ABN66837.
XX
PT New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX
PS Claim 1; Page 3309; 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
SQ Sequence 379 AA;

Query Match 71.9%; Score 1885; DB 5; Length 379;
Best Local Similarity 100.0%; Pred. No. 1.1e-142;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 MTHEHWPVKDLSESELKAAQEFSLGKSEANQDKPKTKGTAQEIYEAIEPKAIVKPEDLL 196
DB 1 MTHEHWPVKDLSESELKAAQEFSLGKSEANQDKPKTKGTAQEIYEAIEPKAIVKPEDLL 60

QY 197 FGIAQATDYKNGTFVIPHKDHVHYVELKWFDEKDLLADSDKTSLEDYLATAKYMMHP 256
DB 61 FGIAQATDYKNGTFVIPHKDHVHYVELKWFDEKDLLADSDKTSLEDYLATAKYMMHP 120

QY 257 EKRPKEGVGKDAEIIYKEKSNKADKPSAPTDNKSTNSSDKNLSAAEVFKQAKPEKIV 316
DB 121 EKRPKEGVGKDAEIIYKEKSNKADKPSAPTDNKSTNSSDKNLSAAEVFKQAKPEKIV 180

QY 317 PLDKIAAHMAYAVGFEDDQILVPHHDHYHVPMAWFDKGLWKAPEGYTLQOLFSTIKYY 376
DB 181 PLDKIAAHMAYAVGFEDDQILVPHHDHYHVPMAWFDKGLWKAPEGYTLQOLFSTIKYY 240

QY 377 MEHPNELPKKEGWGHDSHNKSNKDNKAKNYAPDEEEDSGKVTHNYGYDVNKGSDDEE 436
DB 241 MEHPNELPKKEGWGHDSHNKSNKDNKAKNYAPDEEEDSGKVTHNYGYDVNKGSDDEE 300

QY 437 EPEKQDESELDEYELGMAQNAKKGMDRQSFQKLIQLSNKYSVSFES 485
DB 301 EPEKQDESELDEYELGMAQNAKKGMDRQSFQKLIQLSNKYSVSFES 349

RESULT 3
AAB12756
ID AAB12756 standard; protein; 816 AA.
XX
AC AAB12756;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain P4241 BVH-11-2 protein antigen.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 15:50:08 ; Search time 47 Seconds
(without alignments)
2672.199 Million cell updates/sec

Title: US-09-769-736-72
Perfect score: 2621
Sequence: 1 MEFLAYNAFTAIGVSIIPHGN.....QSFEKQLIQLSNKYSVSFES 485

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2621	100.0	485	10	US-09-769-736-72
2	563	21.5	838	15	US-10-412-850-4
3	563	21.5	838	15	US-10-387-783-4
4	558	21.3	568	15	US-10-324-143-33
5	558	21.3	612	15	US-10-324-143-44
6	558	21.3	819	15	US-10-324-143-32
7	558	21.3	838	10	US-09-884-465A-8
8	558	21.3	838	15	US-10-324-143-9
9	558	21.3	1139	10	US-09-884-465A-380
10	558	21.3	1139	15	US-10-324-143-51
11	558	21.3	1378	10	US-09-884-465A-378
12	558	21.3	1378	15	US-10-324-143-49
13	552	21.1	1238	15	US-10-324-143-52
14	550	21.0	1238	10	US-09-884-465A-381
15	545.5	20.8	819	15	US-10-412-850-10

16	545.5	20.8	819	15	US-10-387-783-10	Sequence 10, Appl
17	545.5	20.8	826	10	US-09-769-787-194	Sequence 194, App
18	543	20.7	555	15	US-10-324-143-26	Sequence 26, Appl
19	543	20.7	613	15	US-10-324-143-22	Sequence 22, Appl
20	543	20.7	821	15	US-10-324-143-19	Sequence 19, Appl
21	543	20.7	840	10	US-09-884-465A-7	Sequence 7, Appli
22	543	20.7	840	15	US-10-324-143-8	Sequence 8, Appli
23	543	20.7	1126	10	US-09-884-465A-383	Sequence 383, App
24	543	20.7	1126	15	US-10-324-143-54	Sequence 54, Appl
25	543	20.7	1365	10	US-09-884-465A-382	Sequence 382, App
26	543	20.7	1365	15	US-10-324-143-53	Sequence 53, Appl
27	542.5	20.7	763	9	US-09-765-272-66	Sequence 66, Appl
28	539.5	20.6	999	10	US-09-884-465A-376	Sequence 376, App
29	537.5	20.5	999	15	US-10-324-143-47	Sequence 47, Appl
30	512.5	19.6	796	9	US-09-765-272-56	Sequence 56, Appl
31	512.5	19.6	819	15	US-10-412-850-8	Sequence 8, Appli
32	512.5	19.6	819	15	US-10-387-783-8	Sequence 8, Appli
33	511.5	19.5	428	15	US-10-324-143-27	Sequence 27, Appl
34	511.5	19.5	999	10	US-09-884-465A-377	Sequence 377, App
35	511.5	19.5	999	15	US-10-324-143-48	Sequence 48, Appl
36	510.5	19.5	342	15	US-10-324-143-40	Sequence 40, Appl
37	510.5	19.5	913	10	US-09-884-465A-384	Sequence 384, App
38	510.5	19.5	913	15	US-10-324-143-55	Sequence 55, Appl
39	510.5	19.5	1152	10	US-09-884-465A-379	Sequence 379, App
40	510.5	19.5	1152	15	US-10-324-143-50	Sequence 50, Appl
41	507.5	19.4	473	15	US-10-324-143-36	Sequence 36, Appl
42	501.5	19.1	487	15	US-10-324-143-21	Sequence 21, Appl
43	501.5	19.1	487	15	US-10-324-143-35	Sequence 35, Appl
44	501.5	19.1	1058	15	US-10-324-143-46	Sequence 46, Appl
45	495.5	18.9	690	15	US-10-324-143-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1
US-09-769-736-72
; Sequence 72, Application US/09769736
; Publication No. US20030138775A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21089wo
; CURRENT APPLICATION NUMBER: US/09/769,736
; CURRENT FILING DATE: 2003-02-14
; PRIOR APPLICATION NUMBER: GB 9816335.5
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125163
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 72
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-09-769-736-72

Query Match 100.0%; Score 2621; DB 10; Length 485;
Best Local Similarity 100.0%; Pred. No. 1.9e-198;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MEFLAYNAFTAIGVSIIPHGNHFHFIHYKDNPSPLELEATRMVAEHRGHHDALGKDKSTEK 60
Db	1	MEFLAYNAFTAIGVSIIPHGNHFHFIHYKDNPSPLELEATRMVAEHRGHHDALGKDKSTEK 60
Qy	61	PKHISHEPNKEPTEEEHHAHVTPKQDKGKPNQIVYSAQIEEAKKAGYTTSDGYIFD 120
Db	61	PKHISHEPNKEPTEEEHHAHVTPKQDKGKPNQIVYSAQIEEAKKAGYTTSDGYIFD 120
Qy	121	AKDIKDDTGTGTGYVIPHMTHEHWVPKDLSESELKAAQEFSLGKSEANQDKPKTKTAQEI 180

Db 121 AKDIKDTGTGYVIPHMTHEHWPVKKOLSESELKAAQEFLSGKSEANQDKPKTKGTAQEI 180
QY 181 YEAIEPKAIVKPEDLLFGIAQAATDYKNGTFVIPHKHGHYHYVELKWFDEEKDLLADSDKTY 240
Db 181 YEAIEPKAIVKPEDLLFGIAQAATDYKNGTFVIPHKHGHYHYVELKWFDEEKDLLADSDKTY 240
QY 241 SLEDYLATAKYMMHPEKRPKVEGWGKDAEIIYKEKDSNKADKPSAPPTDNKSTNSSSDKN 300
Db 241 SLEDYLATAKYMMHPEKRPKVEGWGKDAEIIYKEKDSNKADKPSAPPTDNKSTNSSSDKN 300
QY 301 LSAAEVFKQAKPEKIVPLDKIAAHMAYAVGFEDDQLIVPHHDHYHNVPMWAFDKGGLWKA 360
Db 301 LSAAEVFKQAKPEKIVPLDKIAAHMAYAVGFEDDQLIVPHHDHYHNVPMWAFDKGGLWKA 360
QY 361 PEGYTLQQLFSTIKYYMEHPNELPKEKGWGHDSHNKGSNKNKAKNYAPDEPEDSGKV 420
Db 361 PEGYTLQQLFSTIKYYMEHPNELPKEKGWGHDSHNKGSNKNKAKNYAPDEPEDSGKV 420
QY 421 THNYGFYDVNKGSDDEEPEKQDESELDEYELGMAQNAKKYGMDRQSFQKQLIQLSNKYS 480
Db 421 THNYGFYDVNKGSDDEEPEKQDESELDEYELGMAQNAKKYGMDRQSFQKQLIQLSNKYS 480
QY 481 VSFES 485
Db 481 VSFES 485

RESULT 2
US-10-412-850-4
; Sequence 4, Application US/10412850
; Publication No. US20040001836A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-686
; CURRENT APPLICATION NUMBER: US/10/412,850
; CURRENT FILING DATE: 2003-04-14
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-412-850-4

Query Match 21.5%; Score 563; DB 15; Length 838;
Best Local Similarity 25.0%; Pred. No. 1.3e-35;
Matches 148; Conservative 64; Mismatches 113; Indels 266; Gaps 14;
QY 10 TAIGVSIPIHGNHFIHYKDMSPLELEATRMV-AEHRGHH----- 48
Db 307 TARGVAVPHGNHYHFIPYEQMSELEKRIARIIPLYRSNHWVDSRPEQSPQSTPEPSP 366
QY 49 -----ID-----ALGKKDST 58
Db 367 SPQAPNPQPAPSNPIDEKLVKEAVRKGVDGYVFEENGVSRYIPAKDLSAETAAGIDSKL 426
QY 59 EKPKHISHEPNKE-----PHTEEHH-----A 80
Db 427 AKQESLSHKLGAKTDLPSDDREFYNKAYDILLARIHQDLLDNKGRQVDPEALDNLRLK 486
QY 81 VTPKD-----ORKGKPNQIIVYSAQEIIEEAKKAGKYTTSDGYIFDAK 122
Db 487 DVPSDKVKLVDDILAFAPIRHPERLKGPNQAITYTDDDEIQVAKLAGKYTTEDGYIFDPR 546

QY 123 DIKKTGTGTGYVIPHMTHEHWPVKKOLSESELKAAQEFLSGKSEANQDKPKTKGTAQEIYE 182
Db 547 DITSDEGDAYVTPHMTSHWIKKDSLSEARAAQAAYA----- 584
QY 183 AIEPKAIVKPEDLLFGIAQAATDYKNGTFVIPHKHGHYHYVELKWFDEEKDLLADSDKTYSL 242
Db 585 -----KEKGL----- 589
QY 243 EDYLATAKYMMHPEKRPKVEGWGKDAEIIYKEKDSNKADKPSAPPTDNKSTNSSSDKNLS 302
Db 590 -----TPPSTDHQDSGNTAKGAE 608
QY 303 AAEVFKQAKPEKIVPLDKIAAHMAYAVGFEDDQLIVPHHDHYHNVPMWAFDKGGLWKAPE 362
Db 609 A--IYNRVKAAKKVPLDRMPYNLQYTVVEVKNGSLIPIHYDHYHNIKFEWFDE-CLYEAPK 665
QY 363 GYTLOQLFSTIKYYMEHPNELP-KEKGWGHDSHNKGSNKNKAKNYAPDEE----- 413
Db 666 GYTLEDLLATVKYYVEHPNERPHSDNGFGNASDHVRKNKVDQDSK--PDEKHEHDEVSE 722
QY 414 ---PEDSGKVTH---NYGFYDVNKGSDDEEPEKQDESELDEYELGMAQNA 458
Db 723 PTHPESDEKENHAGLNPSADNLVKPSTDTEETEEAEDTTDEAEIQVENS 773

RESULT 3
US-10-387-783-4
; Sequence 4, Application US/10387783
; Publication No. US20040005331A1
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Polypeptides Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-683
; CURRENT APPLICATION NUMBER: US/10/387,783
; CURRENT FILING DATE: 2003-03-13
; PRIOR APPLICATION NUMBER: 09/468,656
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-387-783-4

Query Match 21.5%; Score 563; DB 15; Length 838;
Best Local Similarity 25.0%; Pred. No. 1.3e-35;
Matches 148; Conservative 64; Mismatches 113; Indels 266; Gaps 14;
QY 10 TAIGVSIPIHGNHFIHYKDMSPLELEATRMV-AEHRGHH----- 48
Db 307 TARGVAVPHGNHYHFIPYEQMSELEKRIARIIPLYRSNHWVDSRPEQSPQSTPEPSP 366
QY 49 -----ID-----ALGKKDST 58
Db 367 SPQAPNPQPAPSNPIDEKLVKEAVRKGVDGYVFEENGVSRYIPAKDLSAETAAGIDSKL 426
QY 59 EKPKHISHEPNKE-----PHTEEHH-----A 80
Db 427 AKQESLSHKLGAKTDLPSDDREFYNKAYDILLARIHQDLLDNKGRQVDPEALDNLRLK 486
QY 81 VTPKD-----ORKGKPNQIIVYSAQEIIEEAKKAGKYTTSDGYIFDAK 122
Db 487 DVPSDKVKLVDDILAFAPIRHPERLKGPNQAITYTDDDEIQVAKLAGKYTTEDGYIFDPR 546
QY 123 DIKKTGTGTGYVIPHMTHEHWPVKKDLSESELKAAQEFLSGKSEANQDKPKTKGTAQEIYE 182
Db 547 DITSDEGDAYVTPHMTSHWIKKDSLSEARAAQAAYA----- 584

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 15:47:27 ; Search time 20 Seconds
(without alignments)
2332.645 Million cell updates/sec

Title: US-09-769-736-72
Perfect score: 2621
Sequence: 1 MEFLAYNAFTAIGVSIPHGN.....QSF EKQLQLSNKYSVSFES 485

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: Pir1:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	563	21.5	839	2 G95115	conserved hypothet
2	546.5	20.9	819	2 B95136	conserved domain p
3	544.5	20.8	853	2 C97985	hypothetical prote
4	542	20.7	855	2 D98004	histidine Motif-Co
5	512.5	19.6	802	2 C95136	conserved domain p
6	512.5	19.6	828	2 E98004	hypothetical prote
7	338	12.9	822	2 T46758	hypothetical 92.4K
8	323	12.3	1039	2 D97985	hypothetical prote
9	321	12.2	1039	2 H95115	conserved hypothet
10	163.5	6.2	667	2 A40713	cyclicin I - bovine
11	151.5	5.8	2364	2 A56577	microtubule-associ
12	149.5	5.7	598	2 B40713	cyclicin I - human
13	149	5.7	182	2 F97985	hypothetical prote
14	146.5	5.6	852	2 A34373	histidine-rich cal
15	145	5.5	4549	2 T20771	hypothetical prote
16	145	5.5	4667	2 T20774	hypothetical prote
17	143.5	5.5	798	2 I50479	neurofilament medi
18	143.5	5.5	1192	2 A71623	probable secreted
19	141.5	5.4	1038	2 JC5497	claustrin - chicke
20	140	5.3	490	2 T43745	clr4 protein - fls
21	140	5.3	1871	2 D96796	probable heat shoc
22	139	5.3	678	2 A54514	glutamic acid-rich
23	137	5.2	490	2 T43700	mating-type loci a
24	137	5.2	2052	2 C97038	phage-related prot
25	136.5	5.2	817	2 S53919	hypothetical prote
26	136	5.2	729	2 S68191	triadin - human
27	136	5.2	2441	2 D71623	erythrocyte membra
28	135	5.2	462	2 G96506	hypothetical prote
29	135	5.2	1115	2 T19137	hypothetical prote

30	134	5.1	522	2 C96608	hypothetical prote
31	133	5.1	2464	1 QRMSPI	microtubule-associ
32	133	5.1	5170	2 T15348	hypothetical prote
33	132.5	5.1	501	2 C71948	hypothetical prote
34	132.5	5.1	622	2 S71342	calnexin precursor
35	132.5	5.1	719	2 A42808	Ig light chain-bin
36	132.5	5.1	1137	2 S13759	morphogenesis-rela
37	131	5.0	713	2 A28706	calpastatin, card
38	130	5.0	699	2 A54660	histidine rich cal
39	129.5	4.9	972	2 S35521	DNA topoisomerase
40	129	4.9	505	2 B64560	poly E-rich protei
41	128.5	4.9	497	2 H70168	hypothetical prote
42	128.5	4.9	535	2 S31097	cold acclimation p
43	128.5	4.9	1500	2 T03824	probable immediate
44	128	4.9	552	2 T25593	hypothetical prote
45	128	4.9	630	2 H89056	protein K09H11.4 [

ALIGNMENTS

RESULT 1

G95115

conserved hypothetical protein SP1003 [imported] - Streptococcus pneumoniae (strain TIGR C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C;Accession: G95115
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: G95115
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-839 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK75120.1; PID:G14972476; GSPDB:GN00164; TIGR:SP4
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP1003
C;Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match	21.5%;	Score 563;	DB 2;	Length 839;
Best Local Similarity	25.0%;	Pred. No. 2.4e-27;		
Matches	148;	Conservative	64;	Mismatches 113; Indels 266; Gaps 14;
Qy	10	TAIGVSIPHGNHFHFIHYKDMSPLELEATRMV-AEHRGH-		48
Db	307	TARGVAVPHGNHYHFIPYEQMSELEKRIARIILPLRYSNHWVPDSRPEQSPQSTPEPSP		366
Qy	49	-----ID-----ALGKDDST		58
Db	367	SPQPAPNPQAPSNPIDEKLVKEAVRKVGDVFEENGVSRYIPAKDLSAETAAGIDSKL		426
Qy	59	EKPKHISHEPNKE-----PHTEEHH-----A		80
Db	427	AKQESLSHKLGAKTDLPSDDREFYNKAYDLLARIHQDLLDNKGRQVDFEALDNLLERLK		486
Qy	81	VTPKD-----QRKGKPNISQIVYSAQIEEAKKAGKYTSDGYIFDAK		122
Db	487	DVPSDKVKLVDDILAFAPIRHPERLGNPAQITTYDDEIQVAKLAGKYTTEDGYIFDPR		546
Qy	123	DIKKDTGTGYVIPHMTHEHWVPKKLSESELKAAQEFLLSGKSEANQDKPKTKTAQEIYE		182
Db	547	DITSDEGDAYVTPHMTSHWIKKDSLSEAAAAQAAYA-----584		
Qy	183	AIEPKAIVKPEDLLFGIAQATDYKNGTFVIPHKDHYYVELKWFDEEKDLADSDKTYSL		242
Db	585	-----KEKGL-----589		
Qy	243	EDYLATAKYMMHPEKRPKVEGWGKDAEIKYKEDSNKADKPSAPPTDNKSTSNSSDKNLS		302

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM protein - protein search, using sw model
Run on: March 23, 2004, 15:43:47 ; Search time 17 Seconds
(without alignments)
1485.531 Million cell updates/sec
Title: US-09-769-736-72
Perfect score: 2621
Sequence: 1 MEFLAYNAFTAIGVSIIPHGN.....OSFEKQLIQLSNKYSVPES 485
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	163.5	6.2	667	1 CYL1 BOVIN	P35662 bos taurus
2	151.5	5.8	2459	1 MAPB_RAT	P15205 rattus norv
3	149.5	5.7	598	1 CYL1 HUMAN	P35663 homo sapien
4	146.5	5.6	852	1 SRCH_RABIT	P16230 oryctolagus
5	139	5.3	678	1 GARP_PLAFF	P13816 plasmodium
6	137	5.2	490	1 CLR4_SCHPO	O60016 schizosacch
7	136.5	5.2	817	1 YG4A_YEAST	P46949 saccharomyc
8	136	5.2	728	1 TRDN_HUMAN	Q13061 homo sapien
9	133.5	5.1	5596	1 MDN1_HUMAN	Q9nu22 homo sapien
10	133	5.1	2464	1 MAPB_MOUSE	P14873 mus musculu
11	132.5	5.1	1137	1 MSB1_YEAST	P21339 saccharomyc
12	131	5.0	713	1 ICAL_PIG	P12675 sus scrofa
13	130	5.0	479	1 T2D6_DROME	Q9vhy5 drosophila
14	130	5.0	699	1 SRCH_HUMAN	P23327 homo sapien
15	130	5.0	700	1 TRDN_CANFA	P82179 canis famil
16	129.5	4.9	972	1 TOP1_DROME	P30189 drosophila
17	128	4.9	644	1 NFM_RABIT	P54938 oryctolagus
18	125	4.8	2116	1 MYS2_DICDI	P08799 dictyosteli
19	124	4.7	1052	1 CLMN_MOUSE	Q8c5w0 mus musculu
20	123	4.7	406	1 NSB1_MOUSE	Q9j135 mus musculu
21	123	4.7	675	1 SGI_RAT	O35314 rattus norv
22	122.5	4.7	1359	1 ATRX_CABEL	Q9u7e0 caenorhabdi
23	121.5	4.6	1220	1 IF2P_HUMAN	O60841 homo sapien
24	121.5	4.6	2300	1 CVAA_NEUCR	Q01631 neurospora
25	120.5	4.6	634	1 SPL1_RAT	P24054 rattus norv
26	120	4.6	810	1 REB1_YEAST	P21538 saccharomyc
27	119	4.5	1453	1 NKCR_MOUSE	P30415 mus musculu
28	118.5	4.5	499	1 VG02_BPP22	P26745 bacterioph
29	118.5	4.5	677	1 UBFI_XENLA	P25979 xenopus lae
30	118.5	4.5	1186	1 CAGA_HELPY	P55980 helicobacte
31	118.5	4.5	1462	1 NKCR_HUMAN	P30414 homo sapien
32	117.5	4.5	500	1 GAR2_SCHPO	P41891 schizosacch
33	117.5	4.5	739	1 VNUC_EBOG4	Q9qce9 ebola virus

34	117	4.5	767	1 TOP1_CERAE	Q7yr26 cercopithec
35	116.5	4.4	401	1 CRTC_EUGGR	Q9zny3 euglena gra
36	116.5	4.4	739	1 VNUC_EBOZ5	O72142 ebola virus
37	116.5	4.4	1085	1 IFH1_YEAST	P39520 saccharomyc
38	116.5	4.4	1483	1 YCDA_DROME	Q9vvgw1 drosophila
39	116.5	4.4	2485	1 PTND_HUMAN	Q12923 homo sapien
40	116.5	4.4	4377	1 ANK3_HUMAN	Q12955 homo sapien
41	116	4.4	716	1 PPPB_BACSU	Q07868 bacillus su
42	116	4.4	793	1 YF06_MYCPN	P75280 mycoplasma
43	116	4.4	915	1 NFM_HUMAN	P07197 homo sapien
44	116	4.4	1087	1 NFM_MOUSE	P19246 mus musculu
45	115.5	4.4	1183	1 CNA_STAAU	Q53654 staphylococ

ALIGNMENTS

RESULT 1
CYL1_BOVIN
ID CYL1_BOVIN STANDARD; PRT; 667 AA.
AC P35662;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Cylicin I (Multiple-band polypeptide I).
GN CYL1 OR CYL.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Testis;
RX MEDLINE=93359502; PubMed=8354692;
RA Hess H., Heid H., Franke W.W.;
RT "Molecular characterization of mammalian cylicin, a basic protein of
the sperm head cytoskeleton.";
RL J. Cell Biol. 122:1043-1052(1993).
CC -!- FUNCTION: Possible architectural role during spermatogenesis. May
be involved in spermatid differentiation.
CC -!- SUBCELLULAR LOCATION: Calyx; sperm head cytoskeletal structure.
CC -!- TISSUE SPECIFICITY: Testis.
CC -!- DEVELOPMENTAL STAGE: Specific to late spermatogenesis.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z22779; CAA80456.1; -.
CC PIR; A40713; A40713.
KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
FT DOMAIN 287 569 9 APPROXIMATE TANDEM REPEATS.
FT REPEAT 287 305 1.
FT REPEAT 306 337 2.
FT REPEAT 338 368 3.
FT REPEAT 369 405 4.
FT REPEAT 406 442 5.
FT REPEAT 443 475 6.
FT REPEAT 476 516 7.
FT REPEAT 517 547 8.
FT REPEAT 548 569 9.
FT DOMAIN 617 667 PRO-RICH.
SQ SEQUENCE 667 AA; 74817 MW; CBF66EA462243D91 CRC64;

Query Match 6.2%; Score 163.5; DB 1; Length 667;
Best Local Similarity 22.6%; Pred. No. 0.027;
Matches 114; Conservative 53; Mismatches 161; Indels 177; Gaps 23;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 15:47:07 ; Search time 46 Seconds
(without alignments)
3326.659 Million cell updates/sec

Title: US-09-769-736-72
Perfect score: 2621
Sequence: 1 MEFLAYNAFTAIGVSIHGN.....QSFKQLIQLSNKYVSFES 485

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTRMBL 25:
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	563	21.5	839	16 Q9ANY2	Q9any2 streptococc
2	547	20.9	481	16 Q8E338	Q8e338 streptococc
3	546.5	20.9	819	16 Q97QM9	Q97qm9 streptococc
4	544.5	20.8	853	16 Q8DQ08	Q8dq08 streptococc
5	542	20.7	844	2 Q9AG74	Q9ag74 streptococc
6	542	20.7	855	16 Q8CWR4	Q8cwr4 streptococc
7	540.5	20.6	819	2 Q9ANY3	Q9any3 streptococc
8	512.5	19.6	802	16 Q97QM8	Q97qm8 streptococc
9	512.5	19.6	816	2 Q9AHT9	Q9aht9 streptococc
10	512.5	19.6	828	16 Q8DPQ2	Q8dpq2 streptococc
11	342.5	13.1	794	16 Q877Y2	Q877y2 streptococc
12	342.5	13.1	823	16 Q8K5Q1	Q8k5q1 streptococc
13	341	13.0	822	16 Q8DZ81	Q8dz81 streptococc
14	340.5	13.0	823	16 Q8NZ82	Q8nz82 streptococc
15	339.5	13.0	825	2 Q93GT5	Q93gt5 streptococc
16	338	12.9	822	2 Q9ZHG7	Q9zhg7 streptococc

17	337.5	12.9	825	16 Q99XV4	Q99xv4 streptococc
18	331	12.6	822	16 Q8E4U1	Q8e4u1 streptococc
19	323	12.3	1039	16 Q8DQ07	Q8dq07 streptococc
20	321	12.2	1039	16 Q9ANY1	Q9any1 streptococc
21	223	8.5	877	16 Q8E5R2	Q8e5r2 streptococc
22	223	8.5	877	16 Q8E029	Q8e029 streptococc
23	208	7.9	289	2 Q9AE21	Q9ae21 streptococc
24	165	6.3	9196	5 Q8IQ87	Q8iq87 drosophila
25	164	6.3	1142	5 Q8T6B4	Q8t6b4 dictyosteli
26	161.5	6.2	958	5 Q8IAW1	Q8iaw1 plasmodium
27	155.5	5.9	792	16 Q8K7L4	Q8k7l4 streptococc
28	153.5	5.9	792	16 Q99Z76	Q99z76 streptococc
29	151	5.8	513	10 Q9LW95	Q9lw95 nicotiana t
30	150.5	5.7	2651	5 Q8IEB6	Q8ieb6 plasmodium
31	150	5.7	637	10 Q9LIQ0	Q9liq0 arabidopsis
32	149.5	5.7	16215	5 Q9NFS3	Q9nfs3 drosophila
33	149	5.7	182	16 Q8DQ06	Q8dq06 streptococc
34	147.5	5.6	792	16 Q8P0G5	Q8p0g5 streptococc
35	147.5	5.6	18074	5 Q9I7U4	Q9i7u4 drosophila
36	147	5.6	556	5 Q9V7I9	Q9v7i9 drosophila
37	146	5.6	542	5 Q8I8F0	Q8i8f0 euplates cr
38	146	5.6	682	5 Q9GTW3	Q9gtw3 plasmodium
39	145	5.5	556	5 Q9S593	Q9s593 drosophila
40	145	5.5	890	10 Q8SAX3	Q8sax3 oryza sativ
41	145	5.5	928	10 Q94GF7	Q94gf7 oryza sativ
42	145	5.5	4541	5 Q7YSV5	Q7ysv5 caenorhabdi
43	145	5.5	4667	5 Q9TVI9	Q9tv19 caenorhabdi
44	144.5	5.5	616	10 Q9FEU6	Q9feu6 populus eur
45	144	5.5	722	6 Q7YS21	Q7ys21 macaca fasc

ALIGNMENTS

RESULT 1

Q9ANY2	PRELIMINARY;	PRT;	839 AA.
ID	Q9ANY2		
AC	Q9ANY2;		
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)		
DE	Pneumococcal histidine triad protein D precursor (Hypothetical protein SP1003) (Fragment).		
GN	PHTD OR SP1003.		
OS	Streptococcus pneumoniae.		
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
OC	Streptococcus.		
OX	NCBI_TaxID=1313;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21101045; PubMed=11159990;		
RA	Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,		
RA	Dormitzer M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,		
RA	Langermann S., Koenig S., Johnson S.;		
RT	"Identification and characterization of a novel family of pneumococcal proteins (the Pht family) that are protective against sepsis.";		
RL	Infect. Immun. 69:949-958(2001).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC BAA-334 / TIGR4;		
RX	MEDLINE=21357209; PubMed=11463916;		
RA	Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,		
RA	Petereson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,		
RA	Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,		
RA	Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,		
RA	Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,		
RA	McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,		
RA	Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,		
RA	Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;		
RT	"Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.";		
RL	Science 293:498-506(2001).		
DR	EMBL; AF318955; AAK06760.1; --		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2004, 15:48:22 ; Search time 23 Seconds
(without alignments)
1088.635 Million cell updates/sec

Title: US-09-769-736-72
Perfect score: 2621
Sequence: 1 MEFLAYNAFTAIGVSIIPHGN.....QSFKEQLIQLSNKYVSFES 485

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	563	21.5	838	4	US-09-468-656A-4	Sequence 4, Appli
2	545.5	20.8	819	4	US-09-468-656A-10	Sequence 10, Appl
3	542.5	20.7	763	3	US-08-961-083-66	Sequence 66, Appl
4	542.5	20.7	763	4	US-09-536-784-66	Sequence 66, Appl
5	512.5	19.6	796	3	US-08-961-083-56	Sequence 56, Appl
6	512.5	19.6	796	4	US-09-536-784-56	Sequence 56, Appl
7	512.5	19.6	819	4	US-09-468-656A-8	Sequence 8, Appli
8	298.5	11.4	484	4	US-09-468-656A-6	Sequence 6, Appli
9	296.5	11.3	447	3	US-08-961-083-182	Sequence 182, App
10	296.5	11.3	447	4	US-09-536-784-182	Sequence 182, App
11	140.5	5.4	1177	4	US-09-134-001C-5106	Sequence 5106, Ap
12	140	5.3	905	3	US-09-074-658-70	Sequence 70, Appl
13	136	5.2	1588	5	PCT-US93-07261-11	Sequence 11, Appl
14	136	5.2	1663	5	PCT-US93-07261-16	Sequence 16, Appl
15	128.5	4.9	535	2	US-08-007-107-2	Sequence 2, Appli
16	126.5	4.8	898	2	US-08-867-941-11	Sequence 11, Appl
17	126.5	4.8	2439	3	US-09-074-658-11	Sequence 11, Appl
18	126	4.8	588	4	US-08-714-741-42	Sequence 42, Appl
19	124.5	4.8	2465	2	US-08-596-291-3	Sequence 3, Appli
20	124.5	4.8	2465	3	US-09-100-804-3	Sequence 3, Appli
21	121.5	4.6	894	2	US-08-867-941-15	Sequence 15, Appl
22	121.5	4.6	2432	3	US-09-074-658-15	Sequence 15, Appl
23	120.5	4.6	864	4	US-08-714-741-40	Sequence 40, Appl
24	120	4.6	1128	4	US-09-641-741-30	Sequence 30, Appl
25	120	4.6	1128	4	US-09-060-482-8	Sequence 8, Appli
26	118	4.5	1128	1	US-08-111-939-2	Sequence 2, Appli
27	116.5	4.4	610	1	US-08-410-804-1	Sequence 1, Appli

28	116.5	4.4	610	1	US-08-259-514-1	Sequence 1, Appli
29	116.5	4.4	610	2	US-08-858-311-1	Sequence 1, Appli
30	116.5	4.4	1085	1	US-08-431-080-28	Sequence 28, Appl
31	116.5	4.4	1085	2	US-08-938-534-28	Sequence 28, Appl
32	116.5	4.4	1085	4	US-09-345-294-28	Sequence 28, Appl
33	116.5	4.4	2466	3	US-09-080-855-12	Sequence 12, Appl
34	116.5	4.4	2466	4	US-09-566-076-12	Sequence 12, Appl
35	116.5	4.4	2466	5	PCT-US94-09943-2	Sequence 2, Appli
36	116.5	4.4	2485	3	US-09-290-640-46	Sequence 46, Appl
37	116.5	4.4	2485	4	US-09-665-615B-46	Sequence 46, Appl
38	116	4.4	693	3	US-08-235-836C-72	Sequence 72, Appl
39	116	4.4	8991	4	US-08-714-741-32	Sequence 32, Appl
40	115.5	4.4	1183	2	US-08-447-031A-2	Sequence 2, Appli
41	115	4.4	660	4	US-09-134-001C-4361	Sequence 4361, Ap
42	115	4.4	1404	4	US-08-801-308-1	Sequence 1, Appli
43	114.5	4.4	809	1	US-07-789-915A-4	Sequence 4, Appli
44	114.5	4.4	809	1	US-08-005-002C-4	Sequence 4, Appli
45	114.5	4.4	809	1	US-08-487-203A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-09-468-656A-4

; Sequence 4, Application US/09468656A

; Patent No. 6582706

; GENERAL INFORMATION:

; APPLICANT: Johnson, Leslie S.

; APPLICANT: Adamou, John E.

; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus

; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural

; TITLE OF INVENTION: Motifs

; FILE REFERENCE: 469201-444

; CURRENT APPLICATION NUMBER: US/09/468,656A

; CURRENT FILING DATE: 1999-12-02

; PRIOR FILING DATE: 1998-12-21

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 838

; TYPE: PRT

; ORGANISM: Streptococcus pneumoniae

US-09-468-656A-4

Query Match 21.5%; Score 563; DB 4; Length 838;
Best Local Similarity 25.0%; Pred. No. 1.6e-37;
Matches 148; Conservative 64; Mismatches 113; Indels 266; Gaps 14;

QY	10	TAIGVSI	PHGNH	FIHYK	DMSPLE	LEATRMV	AEHRGHH	-----	48
Db	307	TARGVAV	PHGNH	YHFI	YEQM	SELEK	RIARI	PLRYR	SNHWV
QY	49	-----	ID	-----	-----	-----	-----	ALGKKDST	58
Db	367	SPQAPNP	QAPSP	NPIDE	KLVKE	AVRKV	GDGV	FEENG	VSRYI
QY	59	EKP	KHIS	HEPN	KE	-----	PHTEEEH	-----	A 80
Db	427	AKQSL	SHKLG	AKKTD	LPSSD	REFY	NKAYD	LLARI	HQDLD
QY	81	VTPKD	-----	-----	-----	-----	-----	-----	-----
Db	487	DVPSD	KVKLV	DDILA	FLAPI	RHPER	LGKPN	AIQIT	YTDDEI
QY	123	DIK	DTGT	GTG	YVIPH	MTHE	HWVP	KKDL	SESEL
Db	547	DITS	DEGD	AYV	TPH	MTSH	WIKD	SLSE	AERAA
QY	183	AIEP	KAIV	KPED	LLFG	IAQAT	DYK	NGT	FVIPH
Db	585	-----	-----	-----	-----	-----	-----	-----	-----

QY 243 EDYLATAKYMMHPEKRPKVEGWGKDAEIIYKEDSNKADKPSAPTNDKSTNSSSDKNLS 302
Db 590 -----TPPSTDHQDSGNTAEKGA 608
QY 303 AAEVFKQAKPEKIVPLDKIAAHMAYAVGFEDDQLIVPHHDHYHNVPMWFDKGLWKAPE 362
Db 609 A--IYNRVKAAKVPDRMPYNLYQTVVEVKNGLIIPHYDHYHNIKFEWFE-GLYEAPK 665
QY 363 GYTLQQLFSTIKYYMEHPNLP-KEKGWGHDSHNGSNKDNKAKNYAPDEE----- 413
Db 666 GYTLLEDLLATVKYYVEHPNERPHSDNGFGNASDHVRKNKVDQDSK---PDEKHEDEVSE 722
QY 414 ---PDSGKVTH---NYGFYDYNKGSDEEEPEKQDESELDYELGMAQNA 458
Db 723 PTHPESDEKENHAGLNPSADNLYKFPSTDTEETEEAEPTTDEAEIPQVENS 773

RESULT 2
US-09-468-656A-10
; Sequence 10, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-10

Query Match 20.8%; Score 545.5; DB 4; Length 819;
Best Local Similarity 24.6%; Pred. No. 4.2e-36;
Matches 151; Conservative 61; Mismatches 106; Indels 295; Gaps 15;
QY 10 TAIGVSIPIHGNHFIHYKDMSPLELEATRMV-AEHRGHH----- 48
Db 306 TARGVAVPHGNHYHFIPIYEQMSELEKRIARIPLRYSNHWVPDSRPEEPSQPTPEPSP 365
QY 49 -----ID-----ALGKDKSTKPKHI 64
Db 366 SPQAPSNPIDGKLVEAVRKVG DG YVFEENGVSRYIPAKDLSAETAAGIDSKLAKQESL 425
QY 55 SHEPNKE---PHTEEHH----- 79
Db 426 SHKLGTKKTDLPSSDRFYFNKAYDILLARIHQDLDNKGQVDFEALDNLRLKDVSSDK 485
QY 80 -----AVTPKDQKGNPSQIVYSAQIEEAKKAGKYTTSDGYIFDAKDIKKDT 128
Db 486 VKLVEDILAFAPIRHPERLGNKPNQIITYTDDIQAQVAKLAGKYTAEDGYIFDPRDITSDE 545
QY 129 GTGYVIPHMTHEHWVPKDLSESELKAAQEFLSGKSEANQDKPKTKGTAQEIYEAIEPKA 188
Db 546 GDAYVTPHMTSHWIKDLSAEARAAQA----- 577
QY 189 IVKPEDLLFGIAQATDYKNGTFVIPHKDHYYVELKNWFEEDKDLADSDKTYSLDYLAT 248
Db 578 -----EKG----- 582
QY 249 AKYMMHPEKRPKVEGWGKDAEIIYKEDSNKADKPSAPTNDKSTNSSSDKNLSAEVFK 308
Db 583 -----TPPSTDHQDSGNTAEKGA---IYN 605

QY 309 QAKPEKIVPLDKIAAHMAYAVGFEDDQLIVPHHDHYHNVPMWFDKGLWKAPEGYTLQ 368
Db 606 RVKAAKVPDRMPYNLYQTVVEVKNGLIIPHYDHYHNIKFEWFE-GLYEAPKGYTLED 664
QY 369 LPSTIKYYMEHPNLP-KEKGWGHDSH---NKGSKND----- 402
Db 665 LLATVKYYVEHPNERPHSDNGFGNASDHVRKNKQADTNQTEKPSSEKPKQTEKPEETP 724
QY 403 -----NKAKNYAPDEEEDSGKVTHNYGFYDYNKGSDEEEPEKQDESELDYELG 453
Db 725 REEPQSEKPESPKPTPEEES-----PEESEEPQVETEKVVEKLEAEADLLG 772
QY 454 MAQ-----NAKK 460
Db 773 KIQDFIISNAKE 785

RESULT 3
US-08-961-083-66
; Sequence 66, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-66

Query Match 20.7%; Score 542.5; DB 3; Length 763;
Best Local Similarity 24.3%; Pred. No. 6.7e-36;
Matches 144; Conservative 59; Mismatches 104; Indels 285; Gaps 13;
QY 10 TAIGVSIPIHGNHFIHYKDMSPLELEATRMV-AEHRGHH----- 48
Db 287 TARGVAVPHGNHYHFIPIYEQMSELEKRIARIPLRYSNHWVPDSRPEEPSQPTPEPSP 346
QY 49 -----ID-----ALGKKOST 58
Db 347 SPQAPNPQAPSNPIDGKLVEAVRKVG DG YVFEENGVSRYIPAKDLSAETAAGIDSKL 406
QY 59 EKPKHISHEPNKE---PHTEEHH----- 79